ChinaXiv合作期刊

第45卷 第6期
2022年11月

干旱区地理
ARID LAND GEOGRAPHY

Vol. 45 No. 6
Nov. 2022

基于Sentinel-2时间序列数据及物候特征的棉花种植区提取

美合日阿依·莫一丁[1]， 买买提·沙吾提[1,2,3]， 李金朝[1]

（1. 新疆大学地理与遥感科学学院，新疆 乌鲁木齐 830046；2. 新疆绿洲生态重点实验室，新疆 乌鲁木齐 840046；3. 智慧城市与环境建模自治区普通高校重点实验室，新疆 乌鲁木齐 840046）

摘 要：棉花是中国重要的经济作物，在新疆大面积种植。及时、准确获取棉花种植面积，对农业政策制定与农业经济发展有重要意义。以渭干河—库车河三角洲绿洲棉花为主要研究对象，利用2018—2020年（1景/1月）36景哨兵2号（Sentinel-2）数据，构建归一化植被指数（Normalize difference vegetation index，NDVI）和红边归一化植被指数（Red edge normalize difference vegetation index，$RENDVI_{783}$）时序数据；采用Savitzky-Golay（S-G）滤波法对时序数据进行平滑、重构并提取11个物候特征；利用袋外误差法对11个物候特征进行特征优选；在此基础上利用重构后的时序数据（NDVI Fit）、（$RENDVI_{783}$ Fit）、物候特征（$RENDVI_{783}$ Ph）、物候特征优选组合构建6种不同的特征数据集，利用随机森林分类（RFC）方法分别进行分类和提取，并采用最大似然分类方法和支持向量机分类方法对分类效果进行验证。结果表明：（1）NDVI和$RENDVI_{783}$时序数据变化趋势较为一致，棉花在5月（苗期）到8月初（开花盛期）有明显的上升趋势，在8月末至9月（花铃期）达到峰值。相比NDVI，红边波段构成的$RENDVI_{783}$时序曲线峰值从0.7提高到0.9，棉花区分效果更佳。（2）11个物候特征中拟合函数最大值、生长季长度、生长季振幅、生长季结束、生长季大积分和生长季小积分对分类的贡献性最大，重要性得分分别为1.43、1.40、1.23、1.16、1.02和1.01。（3）RFC方法对特征数据集（$RENDVI_{783}$ Fit+物候特征优选组合）分类精度最佳。总体精度和Kappa系数分别为92.20%和0.92。（4）研究区内棉花分类精度达到了91.02%，种植面积约为3424 km^2，占研究区总面积的24.67%。

关 键 词：NDVI时间序列；RENDVI时间序列；物候特征；袋外误差；随机森林分类

文章编号：1000－6060(2022)06－1847－13(1847～1859)

棉花作为重要的经济作物，在新疆地区广泛种植。2014年以来，我国在新疆实施棉花目标价格改革，对疆内各地区棉花播种面积、结构、分布产生较大影响[1]。由于新疆环塔里木林果业的迅速发展，果树和棉花等作物种植模式越来越复杂，传统的监测方法无法满足大尺度、复杂地理环境下快速、高精度的棉花识别与面积监测的迫切需求[2]。因此，继续开展棉花高精度识别和信息提取等研究仍有重要意义。

植被物候是指植被发芽、叶片蔓延、抽穗、成熟和脱落的周期性变化，能够反映生长规律，不同作物具有不同的物候特征[3]。因此，可以将这些物候特征进行量化分类。遥感技术的发展为大面积提取农作物物候特征提供了有效的技术途径。国内外很多学者利用多时相Landsat TM[4]、MODIS[5]等中、低分辨率数据的归一化植被指数（Normalize difference vegetation index，NDVI）或增强植被指数（Enhanced vegetation index，EVI）获取作物的生长发育期变化，提取物候信息，进行农作物识别与分类[6-7]。这些影像数据具有高时间分辨率、获取方便与低成本

收稿日期：2022-03-23； **修订日期**：2022-05-16
基金项目：新疆自然科学计划（自然科学基金）联合基金项目（2021D01C055）资助
作者简介：美合日阿依·莫一丁（1997-），女，硕士研究生，主要从事卫星遥感和植被监测等方面的研究. E-mail: mihray_m@163.com
通讯作者：买买提·沙吾提（1976-），男，博士，副教授，主要从事多源遥感信息提取及农业遥感等方面的研究. E-mail: korxat@xju.edu.cn

chinaXiv:202302.00023v1

ChinaXiv合作期刊

等优势应用于地表植被的生长监测和识别,对种植结构单一的作物具有很大的优势[8-9]。但是这些影像数据混合像元量众多,光谱辨识率有限,而且使用的特征变量往往比较单一,因此,在地块破碎、作物生长周期相近、光谱差异小的情况下很难对其进行有效的区分,无法满足大范围、复杂地形条件下的棉花空间分布监测精度要求[10]。

欧洲航天局Sentinel-2卫星,共有13个波段,空间分辨可达10 m,比MODIS、Landsat-8相比在705~865 nm特有3个红边波段[11],为红边波段作物遥感监测提供了数据保障,解决了棉花种植面积监测中精度与成本相互制约的问题。国内外已有很多学者通过Sentinel-2数据探讨了农业监测[12]、土地利用分类[13]、作物分类及面积提取[14]中的应用并证实了红边参数波段对监测植被健康和长势的有效性[15]。

渭干河—库车河三角洲绿洲(简称渭—库绿洲)是塔里木河流域开垦较早并在西北干旱区具有代表性的绿洲。以灌溉农业为主,耕地不断向绿洲外围扩张,种植地块较为细碎。棉花、玉米、果树等生长周期相似的秋季作物套种模式复杂。因此需要开展基于高分辨率数据的棉花识别特征和分类方法的探索。因此,本研究利用36期多时相Sentinel-2数据,在已有NDVI时序数据的遥感监测基础上结合改进的红边归一化植被指数(Red edge normalize difference vegetation index,$RENDVI_{783}$)时序数据,利用Savitzky-Golay (S-G)滤波算法对其进行平滑和重构,提取11个关键性物候特征并其进行特征重要性评价,构建不同的分类特征数据集,用随机森林分类(RFC)方法进行分类和棉花种植面积提取。为农业部门对棉花种植面积监测提供重要的参考依据。

1 研究区概况

渭—库绿洲(82°20′~83°20′E,41°00′~41°40′N)地处天山南麓,塔里木盆地中北部,属于新疆阿克苏地区,包括库车市、沙雅县和新和县[16]。研究区平均海拔971 m,地势北高南低,自西北向东南倾斜,是干旱与极端干旱区典型的扇形平原绿洲。年平均气温10.3 ℃,年降水量50.0~66.5 mm,年均蒸发量2000~2092 mm。作物类型有棉花、玉米、小麦、核桃、红枣等,其中棉花、果树、玉米等秋收作物生长期为4—10月。据统计,渭—库绿洲的棉花面积分别占全疆和阿克苏地区棉花面积的8.56%和38.20%,产量分别占8.41%及40.34%,是新疆棉花主产区[17]。研究区及样点分布如图1所示。

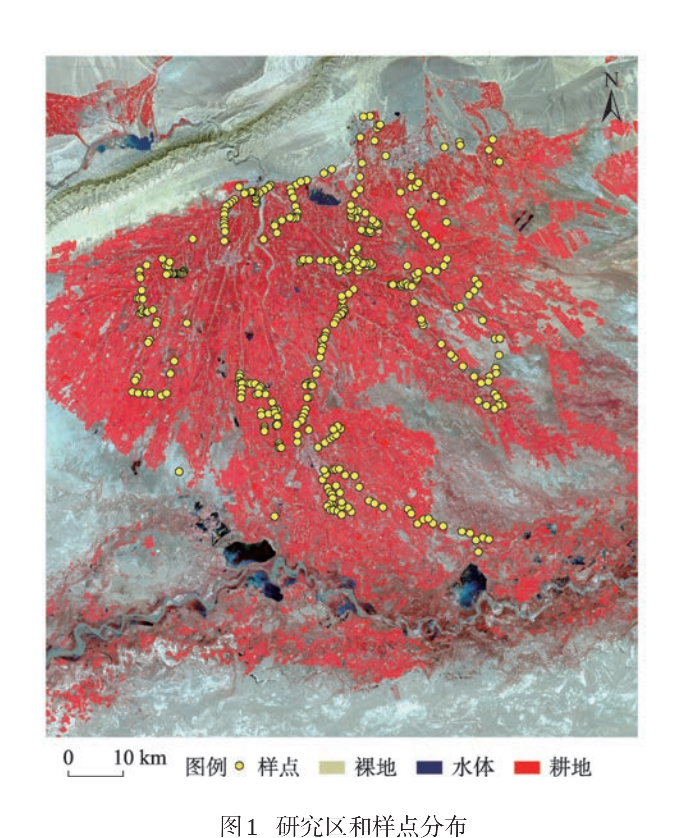


图1 研究区和样点分布

Fig. 1 Schematic diagram of the study area and sample distribution

2 数据与方法

2.1 数据来源与处理

野外实地考察是遥感监测研究的重要组成部分,基于遥感影像分析和解译验证依赖于实地实测数据。实验团队分别于2018年7月、2019年2月、2021年5月进行了野外实地考察,获取了778个典型地物的样点,其中有532个训练样本和246个验证样本,并利用GPS采集每一个样点的经纬度、周围地物分布情况并记录了其他属性,目的是为了考察影像分类所需要的各类地物的样本和不同作物的分布特征。据实地考察和对研究区的概况分类类型最终确定为棉花、玉米、果树、居民地、水体和其他6种地物。

因此,利用2018—2020年作物连续生长季(1景/1月)的36景Sentinel-2数据来构建植被指数时

chinaXiv:202302.00023v1

ChinaXiv合作期刊

chinaXiv:202302.00023v1

序数据，得到每个像元每个时相的植被指数变化信息（表1）。遥感影像云量均在5%以下，部分影像云量超过10%，但这些云主要分布在北部山区，并不影响作物分类。利用SNAP和ENVI等影像处理软件进行预处理，包括大气校正、几何校正、重采样、裁剪等[18]。

表1　遥感影像基本参数

Tab. 1　Basic parameters of remote sensing data

卫星名称	获取时间	影像数量	波段数	空间分辨率/m
	2018年1—12月	12	12	10
Sentinel-2	2019年1—12月	12	12	10
	2020年1—12月	12	12	10

2.2　研究内容

对Sentinel-2时间序列影像进行预处理，构建NDVI和$RENDVI_{783}$时序数据。通过S-G滤波对NDVI、$RENDVI_{783}$时序数据进行平滑、重构并提取11个物候特征，利用袋外误差法（OOB）对物候特征进行特征重要性评价，选取特征重要性得分较高的特征，构建物候特征优选组合。在此基础上利用重构后的时序数据（NDVI Fit和$RENDVI_{783}$ Fit）、物候特征（$RENDVI_{783}$ Ph）、物候特征优选组合等数据集作为分类特征数据集，构建6种不同的分类特征数据集并采用RFC方法进行作物分类及提取。为了验证RFC方法的分类精度，采用最大似然分类（MLC）方法与支持向量机（SVM）分类方法分别对上述6种分类特征数据集进行作物分类及提取。根据分类结果最佳特征组合的分类影像提取研究区棉花及其他作物的面积。

2.3　研究方法

2.3.1　*植被指数时序数据构建*　为了探究棉花在生长过程中的变化，构建两种监测作物生育期和长势规律的指数（NDVI和RENDVI）。将Sentinel-2卫星红波段B4（665 nm）、红边1波段B5（705 nm）、红边2波段B6（740 nm）、红边3波段B7（783 nm）、近红外波段B8（843 nm），按表2公式[19]计算。

2.3.2　*植被指数时序重构及物候特征参数提取*　本研究对时序植被指数进行拟合、平滑、重构，并提取11个物候特征参数，其含义[20]如表3所示。S-G拟合法是由Savitzky和Golay提出的基于最小二乘法的卷积算法，计算公式如下：

表2　植被指数及计算公式

Tab. 2　Indices and calculating formulas of vegetation

特征	指数全称	计算公式
光谱特征	归一化植被指数（NDVI）	(B8-B4)/(B8+B4)
红边特征	红边1归一化植被指数（$RENDVI_{705}$）	(B8-B5)/(B8+B5)
	红边2归一化植被指数（$RENDVI_{740}$）	(B8-B6)/(B8+B6)
	红边3归一化植被指数（$RENDVI_{783}$）	(B8-B7)/(B8+B7)

注：B4、B5、B6、B7、B8分别为红波段（665 nm）、红边1波段（705 nm）、红边2波段（740 nm）、红边3波段（783 nm）、近红外波段（843 nm）。

表3　物候特征

Tab. 3　Phenological characteristics

物候特征	特征含义
生长季开始	左边NDVI增加到用户定义的拟合函数值的时刻
生长季结束	右边NDVI减少到用户定义的拟合函数值的时刻
生长季长度	生长结束与生长季开始的时间差值
生长季基值	拟合函数左右2部分NDVI最小值的平均值
生长季中期	左边增加到80%和右边降低到80%的平均时间
拟合函数最大值	NDVI最大值
生长季振幅	最大值和基值之间均值的差
生长季左导数	NDVI增加到左边振幅80%的时间
生长季右导数	NDVI减少到右边振幅80%的时间
生长季大积分	从生长季开始到生长季结束的区域面积

$$Y_j^* = \sum_{i=-m}^{m} C_i \times Y_{j+1}/N \qquad (1)$$

式中：Y_j^*为平滑后NDVI值；j为原始NDVI时间序列的序数；i为任意一个NDVI时序数据；Y_{j+1}为平滑后的NDVI的序数；C_i为从滤波窗口首部开始第i个NDVI值的权值；N为滤波窗口大小；m为$N/2$滤波窗口长度[21]。

2.3.3　*特征优选方法*　RFC方法不仅可以对遥感影像进行分类，在特征集的选择与优化中发挥着重要作用。从训练样本中随机抽取70%的样本进行分类，假设最初的训练样本集总有N个样本，每个样本都有M个特征。当样本数量较大时，约有30%的样本未被抽取，这一部分被称为OOB，可以估计分类精度，也可以计算参与分类的不同特征的重要性，特征变量重要性评估模型如下[22]：

$$\mathrm{VI}(M_A) = \frac{1}{N}\sum_{t=1}^{N}\left(B_{n_t}^{M_A} - B_{O_t}^{M_A}\right) \qquad (2)$$

式中：VI为特征变量的重要性；M为样本的全部特征数；A为特征值；N为生成的决策树的棵数；t为任意的决策树；O为未加入噪声干扰的时刻；n为加入

ChinaXiv合作期刊

噪声干扰的时刻；$B_{n_t}^{M_A}$ 为任意特征值 M_A 加入噪声干扰时第 t 棵决策树的OOB误差；$B_{O_t}^{M_A}$ 为任意特征值 M_A 未加入噪声干扰时第 t 棵决策树的OOB误差。

2.3.4 分类方法与分类方案 MLC是根据贝叶斯判决准则，在多类判决中用统计的方法建立判别函数集，即将遥感影像多波段数据的分布作为多维正态分布来构造判别分类函数[23]。

$$L(x)=p(C_i|x)=p(x|C_i)\times p(C_i)/p(x) \quad (3)$$

式中：$L(x)$ 为目标函数；x 为事件；i 为任意事件；C_i 为任意事件的类别；$p(C_i|x)$ 为已知 x 事件发生了，那么它属于类别 C_i 的概率；$p(x|C_i)$ 为类别 C_i 对应的 x 的概率；$p(C_i)$ 为类别 C_i 发生的概率；$p(x)$ 为事件 x 发生的概率。

SVM基于通过最小化经验风险和置信区间来最小化结构风险的统计理论。将神经网络结构选择问题转化为比较容易的核函数选择问题，有很好的优势[24]。

$$f(x)=(w\times x)+b\geqslant 1 \quad (4)$$

式中：$f(x)$ 为目标函数；x 为向量之间的内积；w 为权向量；b 为分类阈值，决定超平面距离原点的距离。

RFC方法的主要思路是基于遥感数据及特征变量基础上，首先对数据进行随机重采样，然后构建多个决策树，最后采用多决策树投票的方式确定数据类别归属[25]。

$$\mathrm{Gini}(A)=\sum_{b=1}^{B}p_b(1-p_b) \quad (5)$$

式中：Gini(A)为 A 集合的基尼指数；B 为训练样本中样本种类数；b 为随机选中的样本类别；p_b 为集合 A 中随机选中的样本类别 b 的概率；$(1-p_b)$ 为样本被分错的概率。

表4 用于作物分类的数据组合类型特征

Tab. 4 Combination characteristics of data types used in crop classification

方案	特征集	波段数目
A	重构后的NDVI时序数据(NDVI Fit)	36
B	重构后的 $RENDVI_{783}$ 时序数据($RENDVI_{783}$ Fit)	36
C	重构后的 $RENDVI_{783}$ 11个物候特征($RENDVI_{783}$ Ph)	11
D	物候特征优选组合	6
E	重构后的 $RENDVI_{783}$ 时序数据+11个物候特征($RENDVI_{783}$ Fit+$RENDVI_{783}$ Ph)	47
F	重构后的 $RENDVI_{783}$ 时序数据+物候特征优选组合	42

利用重构后的NDVI时序数据(NDVI Fit)、重构后的 $RENDVI_{783}$ 时序数据($RENDVI_{783}$ Fit)、物候特征($RENDVI_{783}$ Ph)、物候特征优选组合4种数据构建6种不同的分类特征数据集(表4)。

3 结果与分析

3.1 植被指数随时间变化特征

作物的生长规律是播种、出苗、抽雄、成熟和采摘过程[26]，图2表示研究区作物生长周期中植被指数的变化，总体上看4种植被指数呈先上升达到峰值后下降的趋势。从图2a可以看出，3种作物的NDVI曲线较为相似，生长季同时开始，在生长季长度和生长季结束期有较好的区分性。在NDVI时序曲线中，3种作物的峰值均低于0.8，红边波段加入之后，所有作物的RENDVI曲线的峰值均高于0.8，反映出较好的区分效果(图2b~d)。这是因为红边波段(705~783 nm)更深入于作物冠层和叶片，对冠层的微小变化和衰老更加敏感。其中，在 $RENDVI_{783}$ 时间序列中，作物的波峰值较高，约达到0.9。作物曲线之间的分离较好，每种作物的物候规律更加清晰和直观。这是因为构成 $RENDVI_{783}$ 的B7(783 nm)波段与B8波段更接近，从而能够更好地表达植物的物候信息。进一步分析图2d可以看出，棉花的平均生长期约在第180 d左右，NDVI值从第120~140 d苗期开始升高，在第230~250 d花铃期达到峰值，在第270~280 d棉花停止生长时下降，变化特征及其明显。而玉米和果树的生育期为130~150 d左右，第210~240 d达到最大值，易与棉花区分。因此，最终确定利用NDVI和 $RENDVI_{783}$ 时间序列数据进行下一步平滑、重构研究。

3.2 时序数据重构

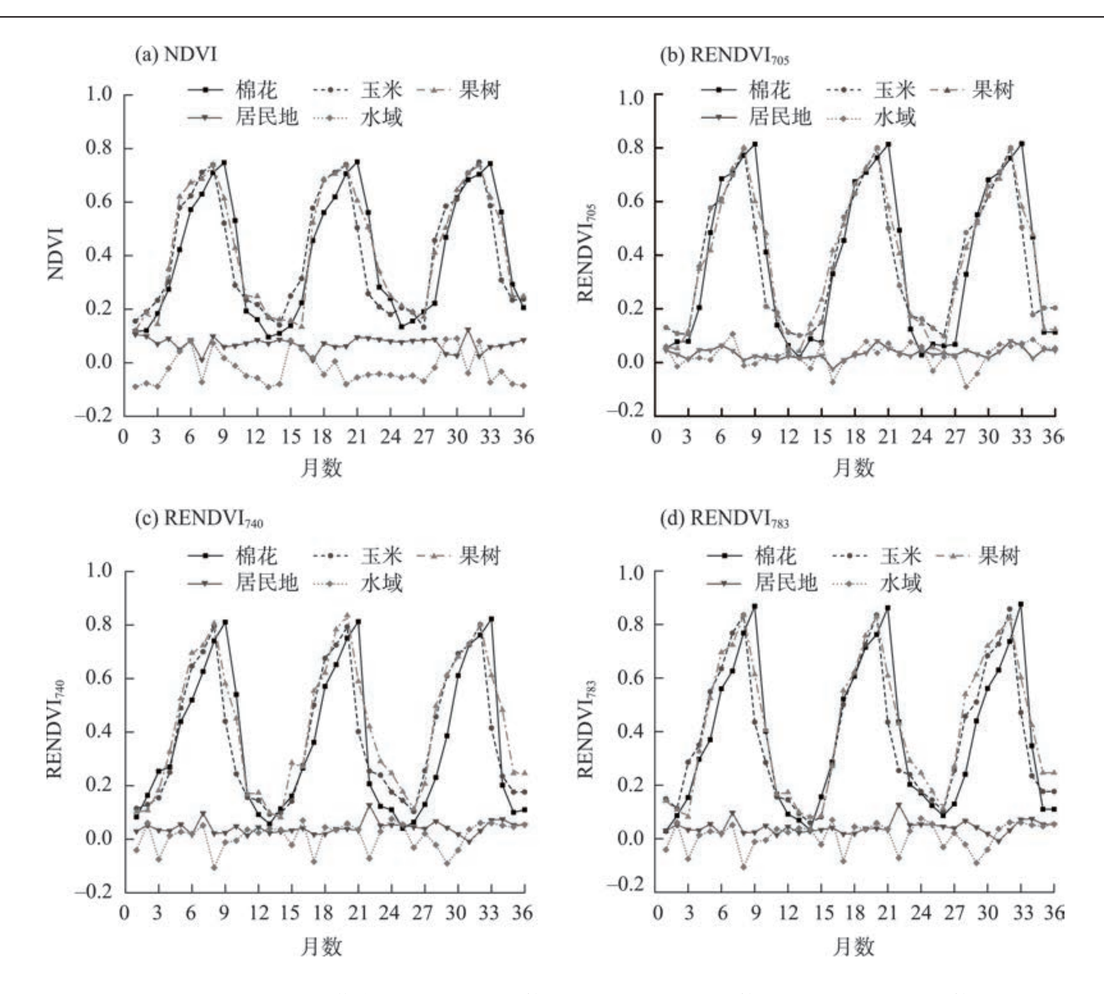
在提取物候特征之前，需要用滤波函数拟合重构植被指数时序影像，最大程度地减少噪声的影响，为后期分类提供准确的数据。图3是纯像元滤波处理前后NDVI和 $RENDVI_{783}$ 时间序列对比曲线。结果表明，经过S-G滤波拟合后，两种时序曲线的噪音和异常值明显降低，曲线更加平滑。平滑后的曲线与研究区3种作物生长周期基本一致，峰值和低谷期与作物的生长盛期及收割期完全吻合。

3.3 物候特征提取

作物的物候特征指标是分类的关键，NDVI和

chinaXiv:202302.00023v1

ChinaXiv合作期刊



注：NDVI、RENDVI705、RENDVI740、RENDVI783依次为归一化植被指数、红边1归一化植被指数、红边2归一化植被指数、红边3归一化植被指数。

图2 不同数据典型地物曲线

Fig. 2 Curves of typical ground object

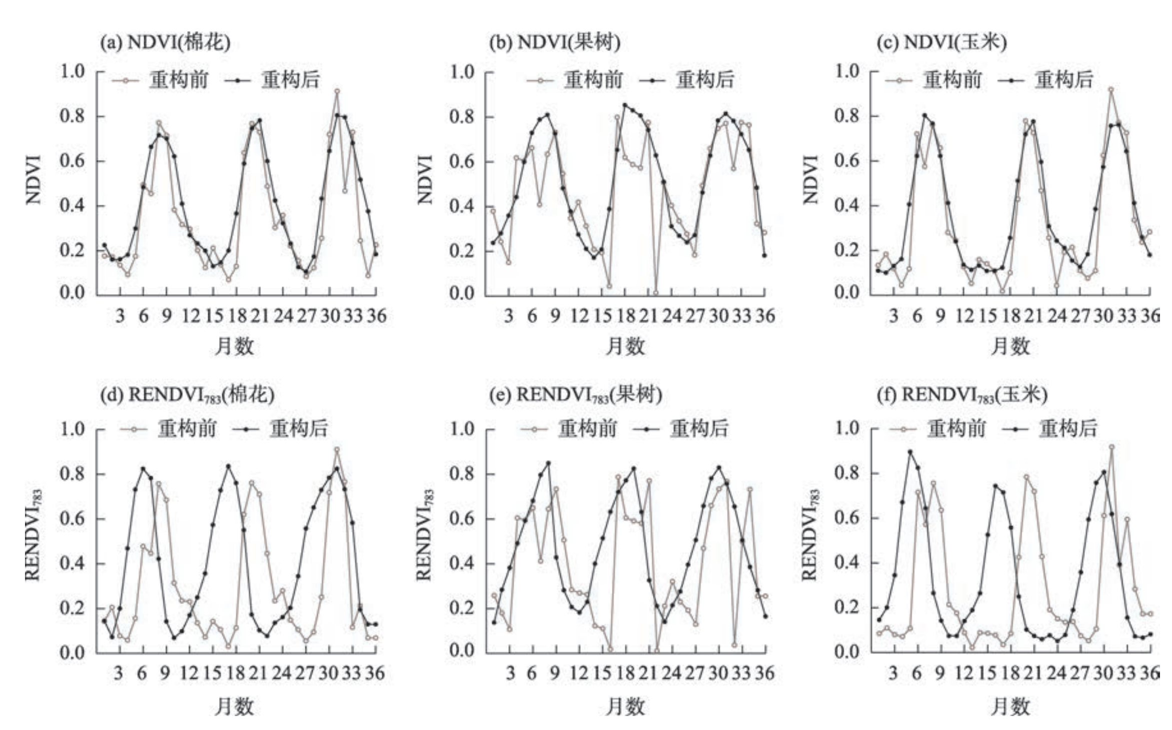
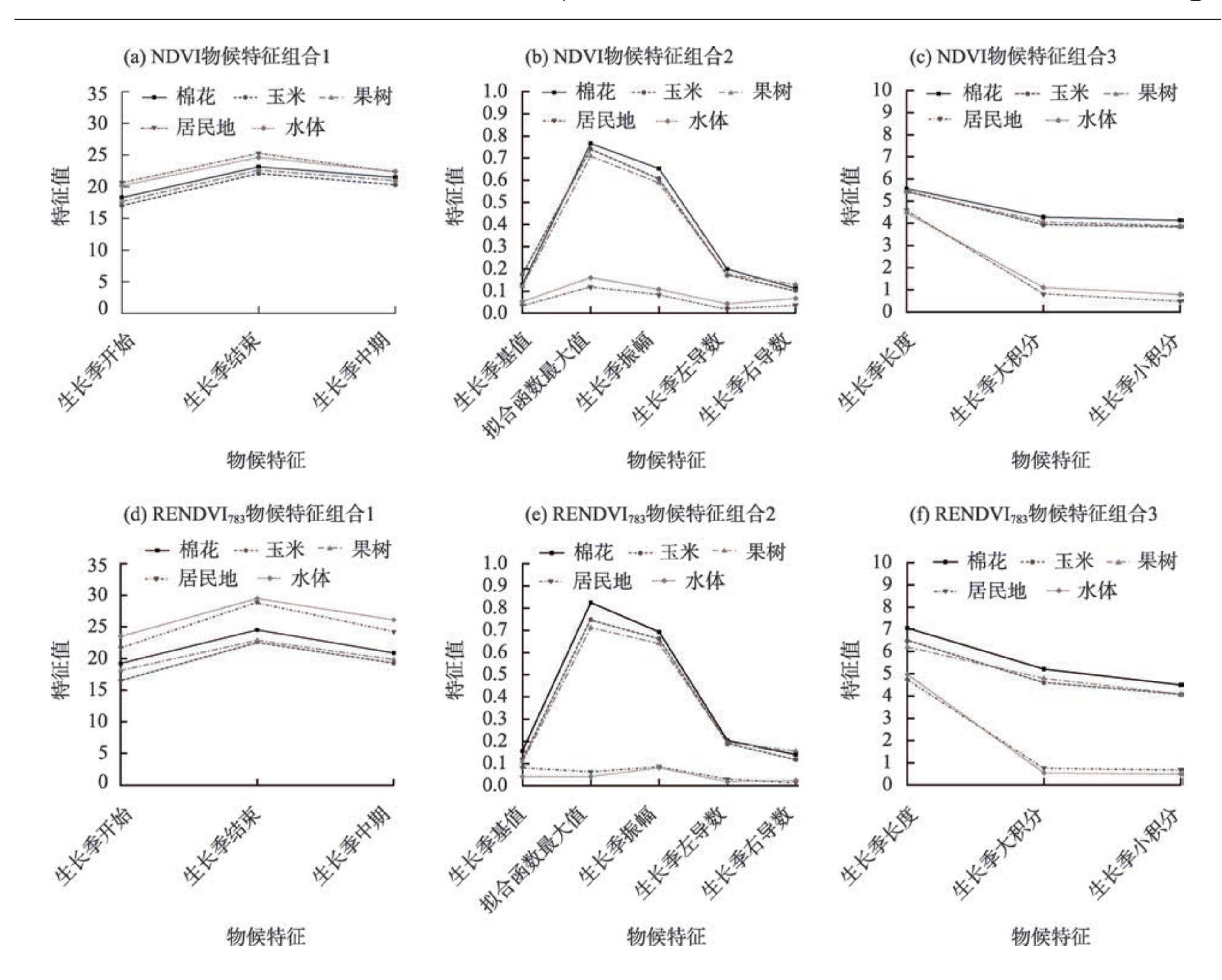


图3 S-G滤波重构前后作物NDVI和RENDVI783时序曲线

Fig. 3 Crop NDVI and RENDVI783 time curves before and after S-G filtering

chinaXiv:202302.00023v1

ChinaXiv合作期刊



注:将11个物候特征按照取值范围分成3组以便进行对比分析。

图4 NDVI和$RENDVI_{783}$同一作物的不同物候特征值

Fig. 4 Different phenological eigenvalues of the same crop in NDVI and $RENDVI_{783}$

$RENDVI_{783}$拟合中每种作物在生长季的物候指标曲线如图4所示。从图4a与图4d生长季开始和生长季结束曲线来看,NDVI比$RENDVI_{783}$曲线起伏程度较小,相对平缓。在生长季开始不同作物在$RENDVI_{783}$上体现的更分散,有区分性。生长季结束指标显示$RENDVI_{783}$出现不同作物集聚为2组。从图4b和4e生长季基值、拟合函数最大值以及生长季振幅3个物候指标值来看,$RENDVI_{783}$比NDVI在不同作物的物候指标有较大差异。拟合函数最大值物候指标显示,$RENDVI_{783}$的波动幅度以及差别要大于NDVI。$RENDVI_{783}$对作物更敏感,棉花与其他作物的差别比较大,最容易识别;从图4c和4f生长季长度、大积分、小积分物候指标中可以看出$RENDVI_{783}$比NDVI波动幅度大,在指标值方面$RENDVI_{783}$比NDVI整体高。生长季长度指标处NDVI出现轻微的集聚,$RENDVI_{783}$相对来说更分散,说明不同作物在生长季长度比较容易区分。因此最终提取重构后$RENDVI_{783}$时序数据的11个物候特征影像并参与到下一步的分类实验。

基于S-G滤波拟合重建后的$RENDVI_{783}$时间序列数据集,选取了11个关键季节性物候特征的指标影像(图5)。可以看出生长季结束期的0值区域波动较少,在拟合函数最大值和生长季振幅的特征值主要集中在研究区北部和渭干河—库车河两岸,在生长季左导数和右导数的0值区域分布较广。

3.4 特征重要性评价

利用RFC方法的OOB法,对上述提取的11个物候特征进行重要性评估,特征的重要性得分如图6。生长季拟合函数最大值的重要性得分最高为1.43,对接下来的分类及信息提取的贡献率最大;其次为生长季长度(1.40);第三为生长季振幅(1.23);生长季结束为1.16;生长季大积分、生长季小积分的

chinaXiv:202302.00023v1

ChinaXiv合作期刊

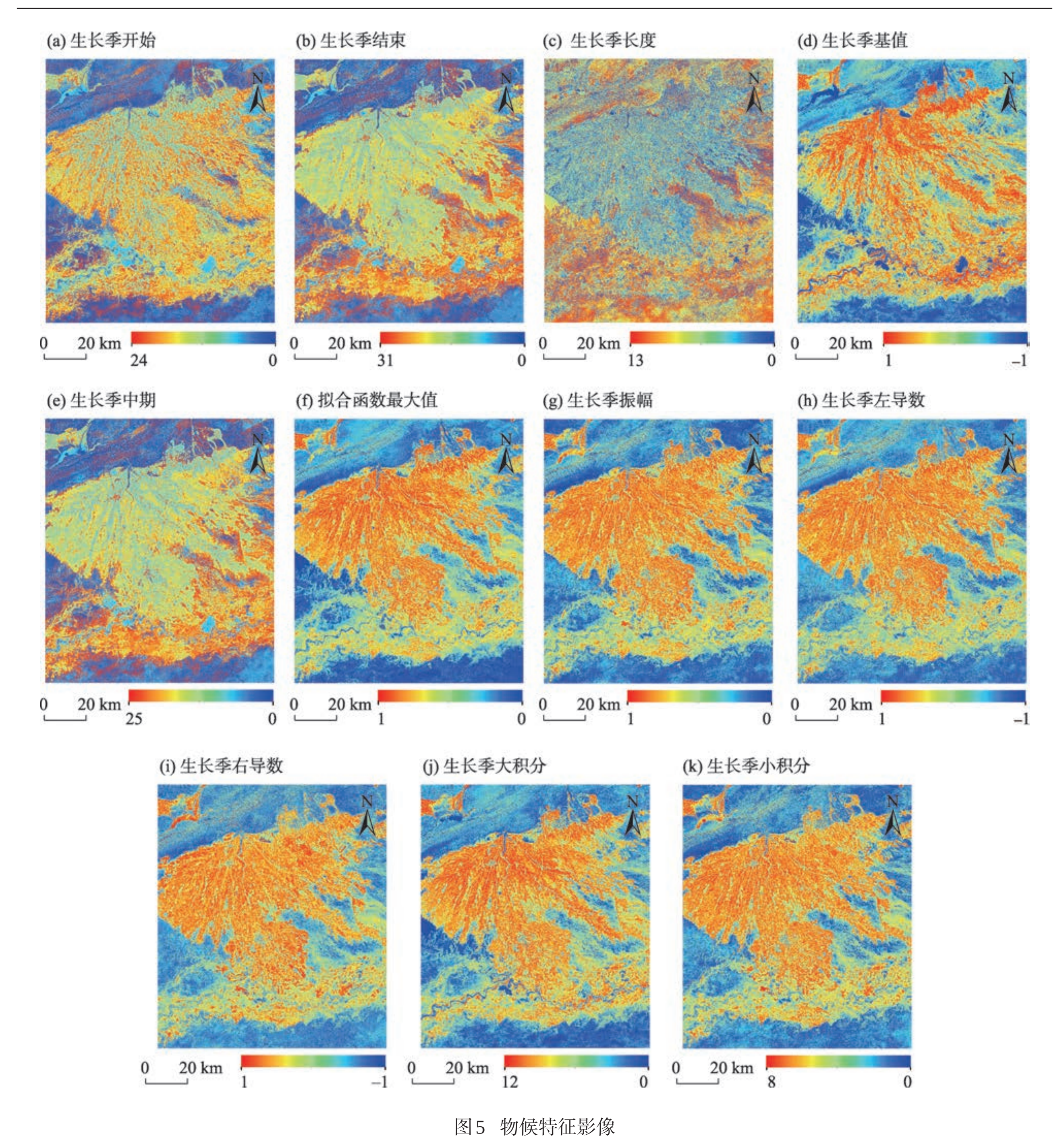


图5 物候特征影像

Fig. 5 Phenological feature image

重要性得分分别为1.02、1.01；其余物候特征的重要性得分均小于1。因此，选取前6个重要性得分较高的特征组成一个优选特征数据集，参与到下一步的分类实验。

3.5 分类结果及精度验证

3种分类方法对6种分类特征方案的分类总体精度从高到低依次为：F方案>C方案>E方案>B方案>D方案>A方案。MLC、SVM和RFC 3种方法中分类效果最好的特征组合均为F方案（$RENDVI_{783}$ Fit+物候特征优选组合），总体精度和Kappa系数分别达到了82.60%、88.40%、92.20%和0.82、0.87、0.92（图7）。

RFC方法中，对比A方案(NDVI Fit)和B方案（$RENDVI_{783}$ Fit），B方案分类精度提高了1.30%，Kappa系数提高了0.02，说明红边波段的参与提高了分类精度；对比C方案（$RENDVI_{783}$ Ph）与E方案（$RENDVI_{783}$ Fit+ $RENDVI_{783}$ Ph），E方案的总体精度和Kappa系数提高了3.40%和0.04，说明物候特征明显提高了分类精度；对比C方案和D方案（物候特征

ChinaXiv合作期刊

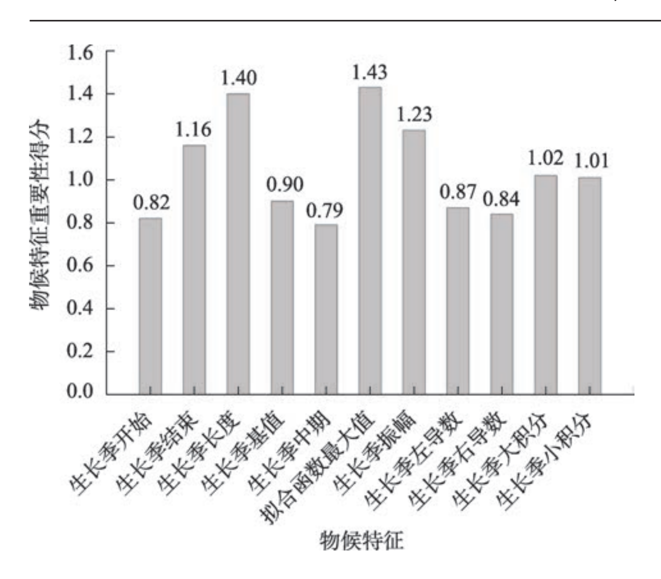


图6 RENDVI783时序物候特征变量的重要性排序

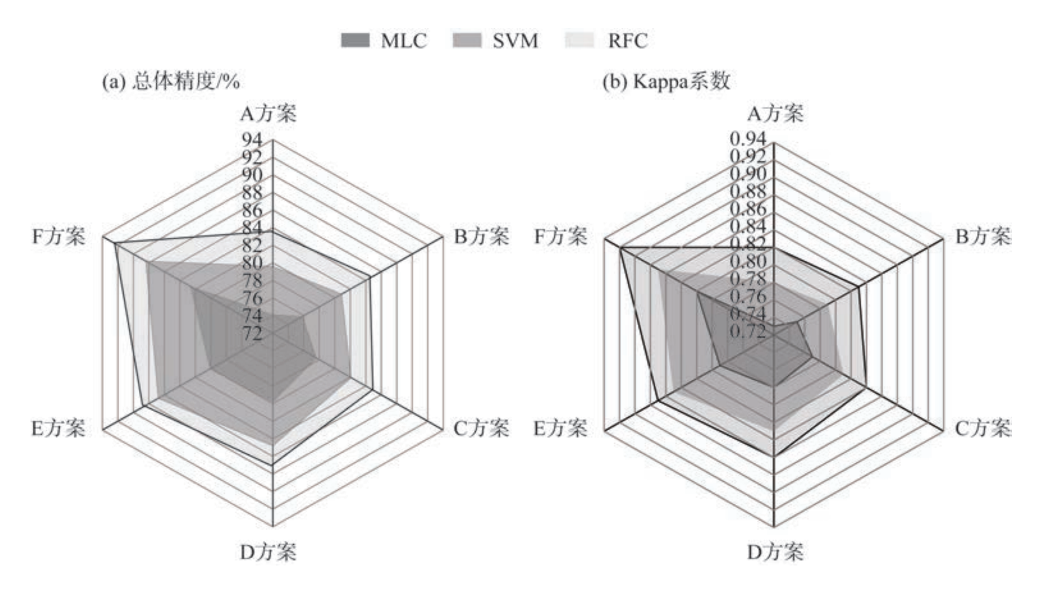
Fig. 6 Importance ordering of feature variables of RENDVI783 time-series phenology

优选组合），D方案的总体精度和Kappa系数提高了2.10%和0.03，说明分类精度不会随着特征波段的增加而增加；对比E方案和F方案，F方案总体精度和Kappa系数提高了3.60%和0.05，关键的特征信息避免交叉冗余信息对棉花提取的干扰，使所有类型的生产者精度和用户精度都有较大程度的改善。

分别对6种特征数据集，3种不同的分类方法所获取的棉花分类结果计算混淆矩阵，具体结果见表5。

MLC方法分类结果从图8a~f中可以看出，图像分类效果整体上比较细碎，棉花的分类效果较差，绿洲边缘地区、东部和西部的棉花种植区域明显稀少。玉米几乎没有被识别出来，大部分玉米误分为居民地。果树沿着渭干河—库车河流域与居民地交错分布，分类面积过多，有误分的情况。

SVM分类结果，从图8g~l中可以看出，没有太



注：MLC为最大似然分类；SVM为支持向量机；RFC为随机森林分类。下同。

图7 3种分类方法不同分类方案精度雷达图

Fig. 7 Precision radar map of three classification methods and different classification plans

表5 3种分类方法不同分类方案棉花分类精度比较

Tab. 5 Comparison of cotton classification accuracy of three classification methods and different plans /%

分类方法	分类精度	A方案	B方案	C方案	D方案	E方案	F方案
MLC	制图精度	79.92	82.06	79.36	82.89	85.05	86.66
	用户精度	71.70	78.39	76.54	78.51	81.92	80.59
SVM	制图精度	83.01	85.03	86.11	90.09	89.33	90.88
	用户精度	76.22	78.21	79.55	82.56	84.17	88.44
RFC	制图精度	87.14	89.07	90.54	91.07	87.09	88.89
	用户精度	79.66	82.32	83.38	85.41	89.36	94.12

注：MLC为最大似然分类；SVM为支持向量机；RFC为随机森林分类。下同。

chinaXiv:202302.00023v1

ChinaXiv合作期刊

chinaXiv:202302.00023v1



图8　不同数据集分类结果对比

Fig. 8　Comparison of different dataset classification results

过细碎的部分，棉花成片状分布在绿洲，“椒盐”现象明显减少。玉米在研究区中部交错分布，果树沿着渭干河—库车河流域与居民地交错分布。研究区的建筑分布较稀疏且散落，东北部分为较大建筑集中区，但是与MLC方法分类相比分类结果比较精确。

RFC分类结果，从图8m~r中可以看出，与上述两种分类结果相比效果比较理想。棉花在研究区由内而外大面积种植，主要在南部和西南部以及塔里木河两岸呈片状分布，因此棉花的识别效果是最好的。有些玉米分布在研究区东部，大部分交错分布在研究区中部。果树大部分沿着2条河流域与居民地交错分布，研究区北部也有大量的种植。由于多种作物套种等复杂的种植方式，使得分类仍然存在错分和漏分现象。该分类结果中，居民地的识别也有显著的提高，主要分布在库车市、新和县和沙雅县人口集中分布的区域，误分情况明显减少。

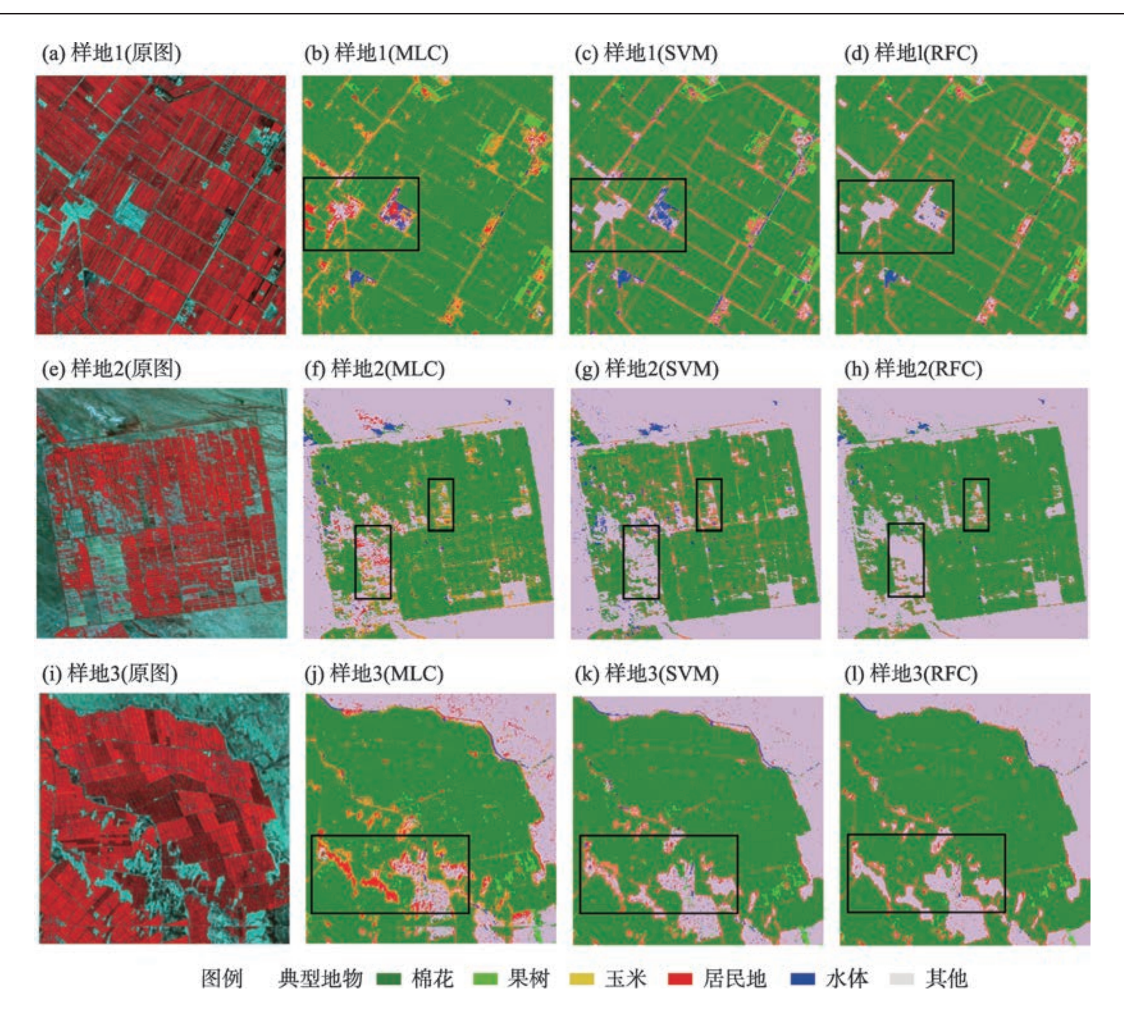
为了突出不同分类方法的分类效果，在3种分类方法中分别选取F方案对应的3个子区域分类结果与原始影像进行对比分析（图9）。样地1的3种方法分类结果局部细节图（图9b~d）可以看出，MLC方法把裸地错分为居民地，SVM分类方法中裸地错分为水体；样地2的3种方法分类结果局部细节图（图9f~h）可以看出，MLC方法裸地错分为玉米，SVM分类方法中裸地错分为居民地，棉花分类面积明显比RFC方法少；样地3的3种方法分类结果局部细节图（图9j~l）可以看出，MLC方法把裸地错分为居民地，SVM分类方法把裸地错分为点状分散地水体，RFC方法效果最好。因此总结出，RFC方法最适合种植模式复杂的研究区，能够有效地识别不同的作物，并得到高精度的提取。

根据RFC方法F方案分类结果，提取研究区棉花、果树、玉米3种作物的面积。研究区主要作物为棉花，种植面积约为3424 km^2，占研究区总面积的24.67%；其次是果园，种植面积约为1216 km^2，占研究区总面积的8.76%；玉米的种植面积最少，约为733 km^2，仅占研究区总面积5.28%。

4　讨 论

本研究分析了Sentinel-2红边归一化植被指数

ChinaXiv合作期刊



注:黑框表示3种分类方法的分类结果出现差异的区域。

图9 3种分类方法的分类结果局部图

Fig. 9 Partial graph of classification results of three different classification methods

chinaXiv:202302.00023v1

时序数据和物候特征对作物分类精度的影响及特征的重要性。利用RFC方法对棉花分布识别与种植面积提取,研究结果在一定程度上解决了以往研究[27]中大尺度遥感影像作物识别过程中混合像元众多和特征变量不足而造成的分类精度较低的问题。同时为时序Sentinel-2数据在干旱区作物的识别及面积提取研究提供了一定的参考。

本研究主要利用新数据源、特征变量的引入和特征优选等方法验证了RFC方法对作物分类和信息提取的可行性。近年来基于物候特征对多时相遥感影像在农作物的信息提取中得到了广泛应用,如MODIS、Landsat等数据构建的NDVI、EVI时序曲线结合作物的季节性特征,较好地提取了甘蔗、水稻等作物的分布和面积信息,精度均比较理想[28-29]。但是对于地块破碎,种植类型比较复杂的区域,Landsat和MODIS数据难以解决混合像元的问题,而且仅靠单一的植被指数难以达到精确分类的目的。本文利用长时间序列Sentinel-2影像不仅提高遥感影像的分辨率,同时为棉花识别提供了敏感波段和有效特征,在时间和空间以及特征量方面扩充作物遥感监测技术与方法体系[30]。另外,本文利用S-G滤波法重建的植被指数时序数据消除原数据存在的噪音和异常值的干扰,提高数据质量的同时保持农作物物候信息的真实性[31]。基于OOB法的物候特征优选避免冗余信息对棉花识别与提取的干扰,使分类精度得到一定程度的提升[32]。最终基于重构后的红边植被指数时序数据和物候特征优选组合的RFC方法精度达到92.20%。研究区棉花种植面积向外延伸,在南部、西南部和塔里木河两岸呈片状分布;但一些套种模式比较复杂的区域存在漏分、误分等现象,作物识别精度有待提高。

虽然本文的研究方法和思路,为克服大范围的干旱区农作物种植面积提取所面临的空间分辨率和参与特征种类不足等问题提供一定的参考,但也

ChinaXiv合作期刊

chinaXiv:202302.00023v1

存在进一步探讨的问题。Sentinel系列卫星的Sentinel-1雷达数据有超强的穿透性、不受云量和天气的影响，与Sentinel-2光学数据具有更好地融合优势。已有学者利用基于多时相双极化SAR数据，根据研究区作物的物候特征和不同作物的生长期后向散射系数，提取了研究区棉花种植面积，分类精度达到了90.88%[33]。鉴于此，以后研究的重点在于利用Sentinel多源遥感数据的综合应用、特征优选方法构建和智能化分类方法建立，从而进一步提高分类精度。

5　结 论

（1）从Sentinel-2 NDVI和$RENDVI_{783}$植被指数的棉花生育期变化规律可以看出，棉花NDVI指数与$RENDVI_{783}$变化趋势较为一致，在5月（苗期）到8月初（开花盛期）有明显的上升趋势，在8月末至9月（花铃期）达到峰值。相比NDVI时序曲线，加入红边波段后的$RENDVI_{783}$时序曲线峰值从0.7提高到0.9，易与相互区分，体现出更佳的区分效果。

（2）利用OOB法对$RENDVI_{783}$的11个物候特征进行重要性评价，生长季拟合函数最大值的重要性最高，为1.43，对分类及信息提取的贡献率最大。其次为生长季长度（1.40）；第三为生长季振幅（1.23）；生长季结束、生长季大积分、生长季小积分的重要性分别为1.16、1.02、1.01。

（3）RFC方法对F方案的分类精度最佳。总体精度和Kappa系数分别为92.20%和0.92。

（4）根据RFC方法F方案的分类结果，提取了研究区内棉花的种植面积，约为3424 km^2，占研究区总面积的24.67%。

参考文献（References）

[1] 王晓伟, 苗红萍, 刘卫东, 等. 目标价格对新疆地方棉区棉花规模的调控效应研究[J]. 中国农业资源与区划, 2017, 38(6): 150-156. [Wang Xiaowei, Miao Hongping, Liu Weidong, et al. The regulatory role of target price on cotton scale in Xinjiang local cotton region[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2017, 38(6): 150-156.]

[2] 王琼, 王克如, 李少昆. HJ卫星数据在棉花种植面积提取中的应用研究[J]. 棉花学报, 2012, 24(6): 503-510. [Wang Qiong, Wang Keru, Li Shaokun. Cotton planting area extraction based on multi-temporal HJ remote sensing[J]. Cotton Science, 2012, 24(6): 503-510.]

[3] 平跃鹏, 臧淑英. 基于MODIS时间序列及物候特征的农作物分类[J]. 自然资源学报, 2016, 31(3): 503-513. [Ping Yuepeng, Zang Shuying. Crop identification based on MODIS NDVI time-series data and phenological characteristics[J]. Journal of Natural Resources, 2016, 31(3): 503-513.]

[4] 熊元康, 张清凌. 基于NDVI时间序列影像的天山北坡经济带农业种植结构提取[J]. 干旱区地理, 2019, 42(5): 1105-1114. [Xiong Yuankang, Zhang Qingling. Cropping structure extraction with NDVI time-series images in the northern Tianshan economic belt[J]. Arid Land Geography, 2019, 42(5): 1105-1114.]

[5] 黄青, 王利民, 滕飞. 利用MODIS_NDVI数据提取新疆棉花播种面积信息及长势监测方法研究[J]. 干旱地区农业研究, 2011, 29(2): 213-217. [Huang Qing, Wang Limin, Teng Fei. MODIS-NDVI-based monitoring of cotton planting areas and growth conditions in Xinjiang[J]. Agricultural Research in the Arid Areas, 2011, 29(2): 213-217.]

[6] 白燕英, 高聚林, 张宝林. 基于Landsat8影像时间序列NDVI的作物种植结构提取[J]. 干旱区地理, 2019, 42(4): 893-901. [Bai Yanying, Gao Julin, Zhang Baolin. Extraction of crop planting structure based on time-series NDVI of Landsat8 images[J]. Arid Land Geography, 2019, 42(4): 893-901.]

[7] 李美丽, 尹礼昌, 张园, 等. 基于MODIS-EVI的西南地区植被覆盖时空变化及驱动因素研究[J]. 生态学报, 2021, 41(3): 1138-1147. [Li Meili, Yin Lichang, Zhang Yuan, et al. Spatio-temporal dynamics of fractional vegetation coverage based on MODIS-EVI and its driving factors in southwest China[J]. Acta Ecologica Sinica, 2021, 41(3): 1138-1147.]

[8] 谢登峰, 张锦水, 潘耀忠, 等. Landsat 8和MODIS融合构建高时空分辨率数据识别秋粮作物[J] .遥感学报, 2015, 19(5): 791-805. [Xie Dengfeng, Zhang Jinshui, Pan Yaozhong, et al. Fusion of MODIS and Landsat 8 images to generate high spatial-temporal resolution data for mapping autumn crop distribution[J]. Journal of Remote Sensing, 2015, 19(5): 791-805.]

[9] 柳文杰, 曾永年, 张猛. 融合时间序列环境卫星数据与物候特征的水稻种植区提取[J]. 遥感学报, 2018, 22(3): 381-391. [Liu Wenjie, Zeng Yongnian, Zhang Meng. Mapping rice paddy distribution by using time series HJ blend data and phenological parameters[J]. Journal of Remote Sensing, 2018, 22(3): 381-391.]

[10] 黄林生, 江静, 黄文江, 等. Sentinel-2影像和BP神经网络结合的小麦条锈病监测方法[J]. 农业工程学报, 2019, 35(17): 178-185. [Huang Linsheng, Jiang Jing, Huang Wenjiang, et al. Wheat yellow rust monitoring method based on Sentinel-2 image and BPNN model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(17): 178-185.]

[11] Sonobe R, Yamaya Y, Tani H, et al. Assessing the suitability of data from Sentinel-1A and 2A for crop classification[J]. Giscience & Remote Sensing, 2017, 54(6): 918-938.

[12] 易秋香. 棉花不同生长阶段Sentinel-2卫星植被指数变化特征[J]. 中国棉花, 2019, 46(8): 13-18. [Yi Qiuxiang. The change fea-

ChinaXiv合作期刊

tures of Sentinel-2 vegetation indices with cotton growth[J]. China Cotton, 2019, 46(8): 13-18.]
[13] 张磊, 宫兆宁, 王启为, 等. Sentinel-2影像多特征优选的黄河三角洲湿地信息提取[J]. 遥感学报, 2019, 23(2): 313-326. [Zhang Lei, Gong Zhaoning, Wang Qiwei, et al. Wetland mapping of Yellow River Delta wetlands based on multi-feature optimization of Sentinel-2 images[J]. Journal of Remote Sensing, 2019, 23(2): 313-326.]
[14] 何云, 黄翀, 李贺, 等. 基于Sentinel-2A影像特征优选的随机森林土地覆盖分类[J]. 资源科学, 2019, 41(5): 992-1001. [He Yun, Huang Chong, Li He, et al. Land-cover classification of random forest based on Sentinel-2A image feature optimization[J]. Resources Science, 2019, 41(5): 992-1001.]
[15] 马依拉·热合曼, 买买提·沙吾提, 尼格拉·塔什甫拉提, 等. 基于遥感与GIS的渭—库绿洲生态系统服务价值时空变化研究[J]. 生态学报, 2018, 38(16): 5938-5951. [Reheman Mayila, Sawuti Mamat, Taxipulati Nigela, et al. The ecosystem service value spatial-temporal changes in the Ugan-Kuqa River Delta Oasis based on RS and GIS [J]. Acta Ecologica Sinica, 2018, 38(16): 5938-5951.]
[16] 安赛日丁·艾尔肯, 买买提·沙吾提, 李金朝. 基于冠层光谱特征的长绒棉叶绿素含量估算[J]. 激光与光电子学进展, 2022, 59 (5): 1-11. [Arkin Ansardin, Sawut Mamat, Li Jinzhao. Estimation of chlorophyll content of long-staple cotton based on canopy spectrum characteristics[J]. Laser & Optoelectronics Progress, 2022, 59(5): 1-11.]
[17] 曹霖. 基于Sentinel-2影像的延庆区森林蓄积量遥感估测研究[D]. 北京: 北京林业大学, 2019. [Cao Lin. Estimation of forest stock volume in Yanqing District based on Sentinel-2 image[D]. Beijing: Beijing Forestry University, 2019.]
[18] Sharma L K, Bu H G, Denton A, et al. Active-optical sensors using red NDVI compared to red edge NDVI for prediction of corn grain yield in north Dakota, U.S.A[J]. Sensors, 2015, 15(11): 27832-27853.
[19] 刘德强. 基于MODIS时间序列数据的广西壮族自治区作物分类研究[D]. 南宁: 广西师范学院, 2017. [Liu Deqiang. Study on crop classification of Guangxi Zhuang Autonomous Region on MODIS time series data[D]. Nanning: Guangxi Teacher Education University, 2017.]
[20] 宋春桥, 柯灵红, 游松财, 等. 基于TIMESAT的3种时序NDVI拟合方法比较研究——以藏北草地为例[J]. 遥感技术与应用, 2011, 26(2): 147-155. [Song Chunqiao, Ke Linghong, You Songcai, et al. Comparison of three NDVI time-series fitting methods based on TIMESAT: Taking the grass land in northern Tibet as case[J]. Remote Sensing Technology and Application, 2011, 26(2): 147-155.]
[21] 孟祥珍, 刘丹, 黄可, 等. 基于植被物候特征的互花米草提取方法研究——以长三角湿地为例[J]. 海洋通报, 2021, 40(5): 591-600. [Meng Xiangzhen, Liu Dan, Huang Ke, et al. Extraction method of *Spartina alterniflora* based on vegetation phenology characteristics: A case study of wetlands in the Changjiang River Delta [J]. Marine Science Bulletin, 2021, 40(5): 591-600.]
[22] 张磊, 宫兆宁, 王启为, 等. Sentinel-2影像多特征优选的黄河三角洲湿地信息提取[J]. 遥感学报, 2019, 23(2): 313-326. [Zhang Lei, Gong Zhaoning, Wang Qiwei, et al. Wetland mapping of Yellow River Delta wetlands based on multi-feature optimization of Sentinel-2 images[J]. Journal of Remote Sensing, 2019, 23(2): 313-326.]
[23] 孙琳, 宋爱红, 高文秀. 基于最大似然法和SVM法的太湖流域HJ-1B影像分类[J]. 测绘信息与工程, 2012, 37(3): 30-33. [Sun Lin, Song Aihong, Gao Wenxiu. Land use/cover classification in Taihu Lake Basin with HJ-1B image: Comparing maximum likelihood classification and support vector machine methods[J]. Journal of Geomatics, 2012, 37(3): 30-33.]
[24] Pal M, Foody G M. Feature selection for classification of hyperspectral data by SVM[J]. IEEE Transactions on Geoscience & Remote Sensing, 2010, 48(5): 2297-2307.
[25] 王利民, 刘佳, 杨玲波, 等. 随机森林方法在玉米-大豆精细识别中的应用[J]. 作物学报, 2018, 44(4): 569-580. [Wang Limin, Liu Jia, Yang Lingbo, et al. Application of random forest method in maize-soybean accurate identification[J]. Acta Agronomica Sinica, 2018, 44(4): 569-580.]
[26] 郝鹏宇. 基于多时相遥感数据的农作物分类研究[D]. 北京: 中国科学院大学, 2017. [Hao Pengyu. Crop classification using time-series remote sensing data[D]. Beijing: University of Chinese Academy of Sciences, 2017.]
[27] 伊尔潘·安尼瓦尔, 买买提·沙吾提, 买合木提·巴拉提. 基于GF-2影像和Unet模型的棉花分布识别[J/OL]. 自然资源遥感. [2022-04-26]. http://kns.cnki.net/kcms/detail/10.1759.P.20211109.1111.006.html. [Anwar Erpan, Sawut Mamat, Balati Maihemuti. Cotton distribution recognition based on GF-2 image and Unet medel[J/OL]. Remote Sensing for Natural Resources. [2022-04-26]. http://kns.cnki.net/kcms/detail/10.1759.P.20211109.1111.006.html.]
[28] 韩振锋. 融合环境卫星和MODIS数据的甘蔗种植区提取[D]. 南宁: 南宁师范大学, 2019. [Han Zhenfeng. Sugarcane extraction based on fused HJ satellite and MODIS data[D]. Nanning: Nanning Normal University, 2019.]
[29] 陈波, 王加胜. 基于Landsat影像的马龙区水稻种植区遥感提取与分析[J]. 江苏科技信息, 2021, 38(5): 56-58. [Chen Bo, Wang Jiasheng. Remote sensing extraction and analysis of rice planting areas in Malong District based on Landsat images[J]. 2021, 38(5): 56-58.]
[30] 彭壮壮, 武永斌, 朱梦豪, 等. 基于时序Sentinel-2影像的农作物分类研究[J]. 测绘与空间地理信息, 2021, 44(12): 81-84. [Peng Zhuangzhuang, Wu Yongbin, Zhu Menghao, et al. Research on crop classification based on time series Sentinel-2 images[J]. Geomatics & Spatial Information Technology, 2021, 44(12): 81-84.]
[31] 周增光, 唐娉. 基于质量权重的Savitzky-Golay时间序列滤波方法[J]. 遥感技术与应用, 2013, 28(2): 232-239. [Zhou Zengguang, Tang Ping. VI-quality-based Savitzky-Golay method for time series data[J]. Remote Sensing Technology and Application, 2013, 28(2):

chinaXiv:202302.00023v1

ChinaXiv合作期刊

232−239.]

[32] 姚茂林, 江洪, 张丽玉. 基于Sentinel−2影像红边光谱指数与特征优选的竹林提取研究[J/OL]. 海南大学学报(自然科学版). [2022−04−25]. http://kns.cnki.net/kcms/detail/46.1013.N.20220421.1203.002.html. [Yao Maolin, Jiang Hong, Zhang Liyu. Bamboo extraction using improved spectral indices and random forest gini index [J/OL]. Natural Science Journal of Hainan University. [2022−04−25]. http://kns.cnki.net/kcms/detail/46.1013.N.20220421.1203.002.html.]

[33] 古丽努尔·依沙克, 买买提·沙吾提, 马春玥. 基于多时相双极化SAR数据的作物种植面积提取[J]. 作物学报, 2020, 46(7): 1099−1111. [Ishake Gulnur, Shawuti Mamat, Ma Chunyue. Extraction of crop acreage based on multi-temporal and dual-polarization SAR data[J]. Acta Agronomica Sinica, 2020, 46(7): 1099−1111.]

Extraction of cotton planting area based on Sentinel−2 time series data and phenological characteristics

Mihray MOYIDIN[1], Mamat SAWUT[1,2,3], LI Jinzhao[1]

(1. College of Geography and Remote Sensing Science, Xinjiang University, Urumqi 830046, Xinjiang, China; 2. Xinjiang Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi 830046, Xinjiang, China; 3. Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, Xinjiang University, Urumqi 830046, Xinjiang, China)

Abstract: Cotton is an important economic crop in China and is cultivated on a large scale in Xinjiang. The timely and accurate acquisition of cotton planting areas is significant to agricultural policy formulation and economic development. This study takes the Ugan River-Kuqa River delta oasis cotton as the main research object and uses the Sentinel−2 data of 36 scenes from 2018 to 2020 (1 scene/month) to construct a normalized vegetation index (NDVI) and red-edge normalized vegetation index (RENDVI) time-series data; The Savitzky-Golay filtering method is used to smooth and reconstruct the time-series data, and 11 phenological features are extracted. The out-of-bag method is used to optimize the 11 phenological features. Time-series data NDVI Fit and $RENDVI_{783}$ Fit, phenological feature $RENDVI_{783}$ Ph, and phenological feature optimization were combined to construct six different feature datasets, which were classified and extracted by random forest classification (RFC). The maximum likelihood and support vector machine were used to verify the classification effect. The results are as follows: (1) The changing trends of NDVI and $RENDVI_{783}$ time-series data are relatively consistent. Cotton has an obvious upward trend from May (seedling stage) to early August (blooming stage) and reaches a peak from the end of August to September (flowering and boll stage). Compared with NDVI, the peak value of the $RENDVI_{783}$ time-series curve composed of the red-edge band is increased from 0.7 to 0.9, and the cotton discrimination effect is better. (2) Among the 11 phenological features, the maximum value of the fitting function, the length of the growing season, the amplitude of the growing season, the end of the growing season, the large integral of the growing season, and the small integral of the growing season contributed the most to the classification effect, with importance scores of 1.43, 1.40, 1.23, 1.16, 1.02, and 1.01, respectively. (3) The RFC method has the best classification accuracy for the feature dataset ($RENDVI_{783}$ Fit + phenological feature optimal combination). The overall accuracy and Kappa coefficient are 92.20% and 0.92, respectively. (4) The cotton classification accuracy in the study area reached 91.02%, and the planting area was approximately 3424 km^2, accounting for 24.67% of the total area of the study area.

Key words: NDVI time-series; RENDVI time-series; phenological characteristics; out-of-bag error method; random forest classification

chinaXiv:202302.00023v1